弱教師学習手法を用いたWebからの食事検出器の自動学習

PRMU 2016

下田　和・柳井啓司（電通大）

本研究の目的

- 弱教師あり学習による食事の検出
- シングルラベルのみを学習に用いて、複数の食事を検出

学習画像

ごはん

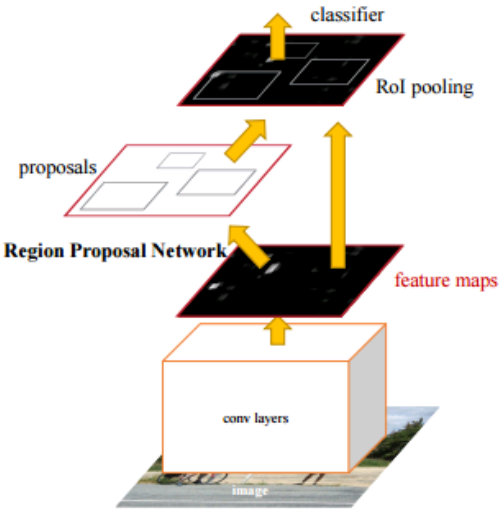
おでん

味噌汁

テスト画像

完全教師あり物体検出

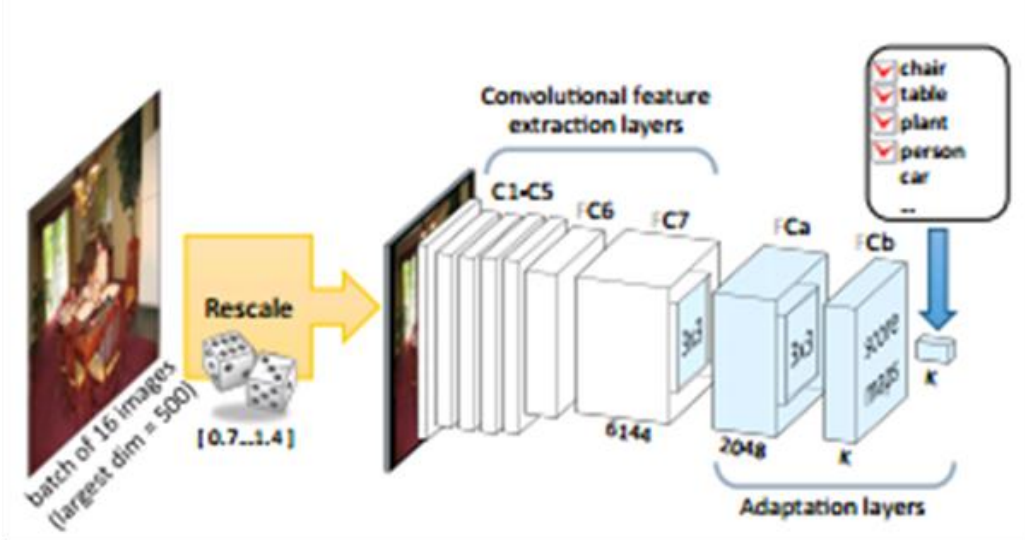
- Faster RCNN
  - バウンディングボックスのアノテーションを用いて学習
  - 物体検出において高精度を達成



[Ren et al. NIPS 2015]

弱教師あり物体検出

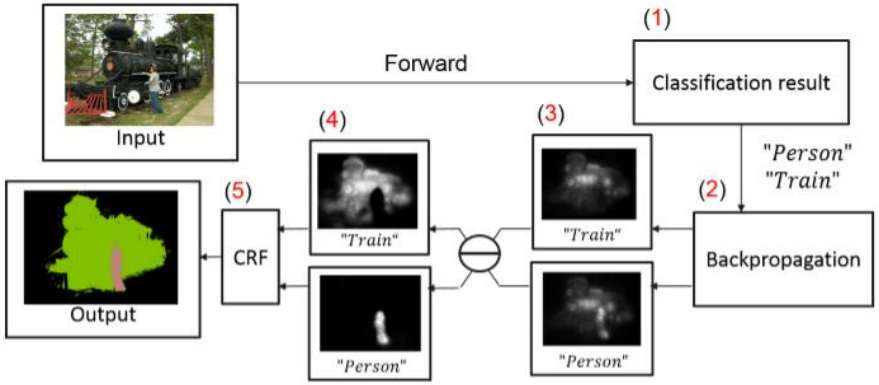
- Fully Convolutional Network + Global Max Pooling
  - シングルラベル＋マルチラベルから学習
  - マルチクラス分類、物体検出で高精度を達成

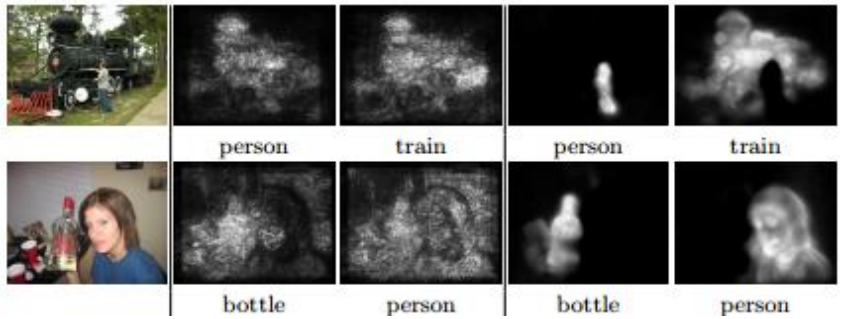


[Oquab et al. CVPR 2015]

弱教師あり領域分割

- 逆伝搬値の差分による領域分割
  - Simonyan らの改良手法
  - シングルラベル＋マルチラベルから学習
  - 弱教師あり領域分割で高精度を達成





[Shimoda et al. ECCV 2016]

本研究

- シングルラベルのみを学習に用いた弱教師あり物体検出
  - 既存の弱教師ありによる位置推定手法は、マルチラベルを含むデータセット（Pascal VOC、MSCOCO）を用いた学習を想定
  - Web画像などから得られる画像の多くはシングルラベル
  - シングルラベルのみを学習に用いて、複数物体を検出する学習手法は多く研究されていない

シングルラベルとマルチラベルの違い

- 学習画像
  - シングルラベル

刺身

ゴーヤーチャンプルー

冷ややっこ

- テスト画像
  - マルチラベル

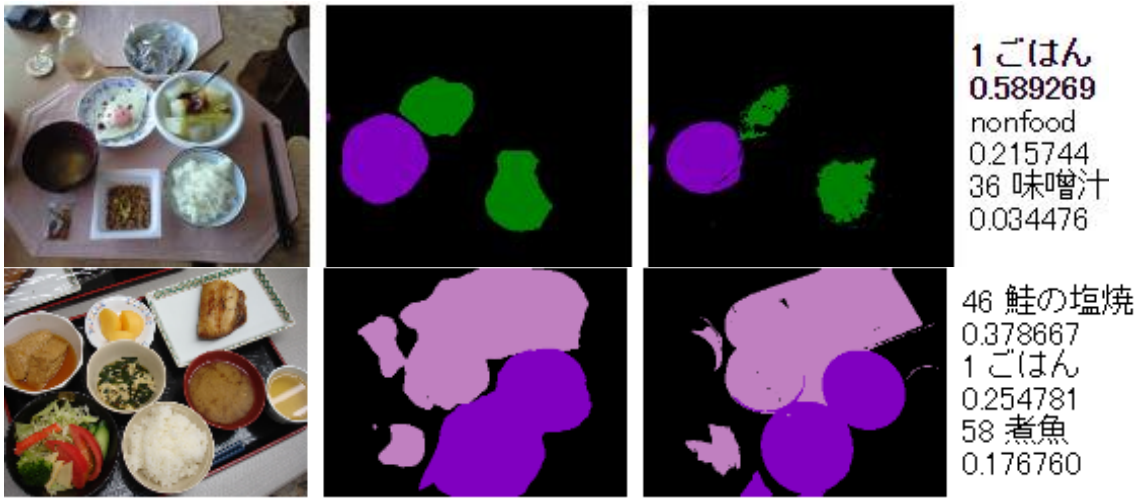
サラダ
ごはん
味噌汁
焼き魚

ひじき
卵焼き
ごはん
味噌汁

筑前煮
ごはん
味噌汁
漬物

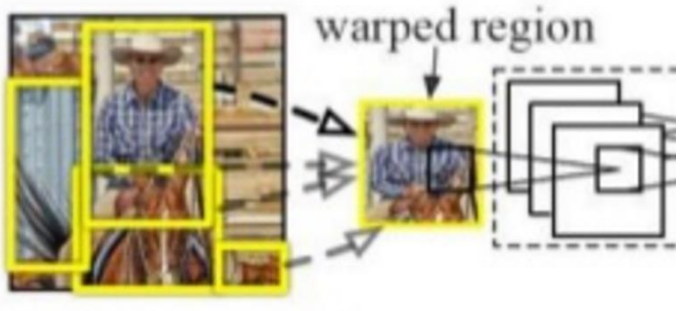
既存のマルチラベルを用いる弱教師あり領域分割手法

- 逆伝搬値の差分による領域分割
  - 低精度な位置推定結果
  - 共起情報による認識過程の変化
  - 学習データにない対象を認識できない



研究の方針

- プロポーザルを用いた物体検出・領域分割
  - シングルクラスの画像で学習した識別器を適用可能
  - ボトムアップ
  - 既存研究　RCNN
    - 約2000の低精度なプロポーザル
    - 計算コストが大きい

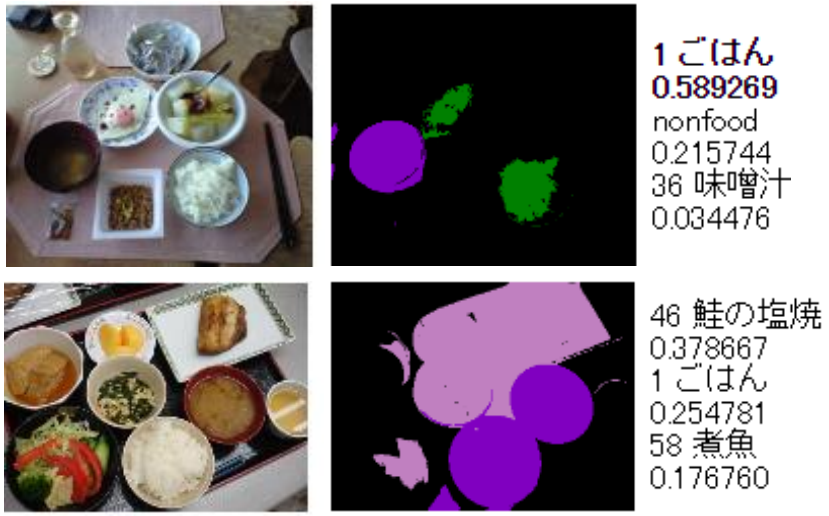


[Girshk et al. CVPR 2014]

既存のマルチラベルを用いる弱教師あり領域分割手法

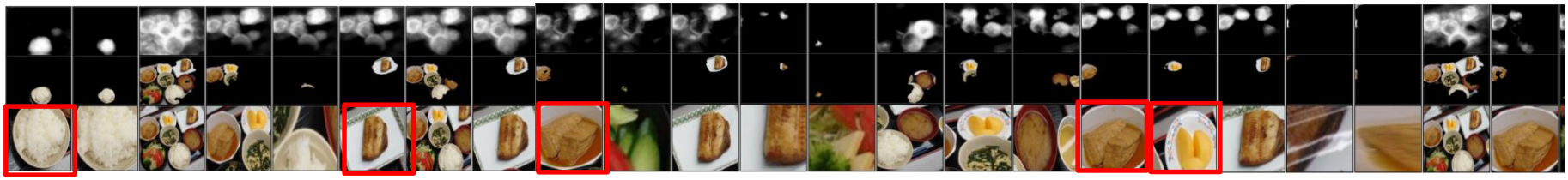
- 逆伝搬値の差分による領域分割
  - 低精度な位置推定結果
  - 位置推定結果は食事領域に帰属
  - 大まかな食事領域の位置推定

食事領域の
プロポーザルへ



領域候補の生成

- 上位Nクラスの位置推定結果をプロポーザルとして扱う
  - 実際には画像に存在しない食事クラスの位置推定結果も何らかの食事に反応



領域候補の生成

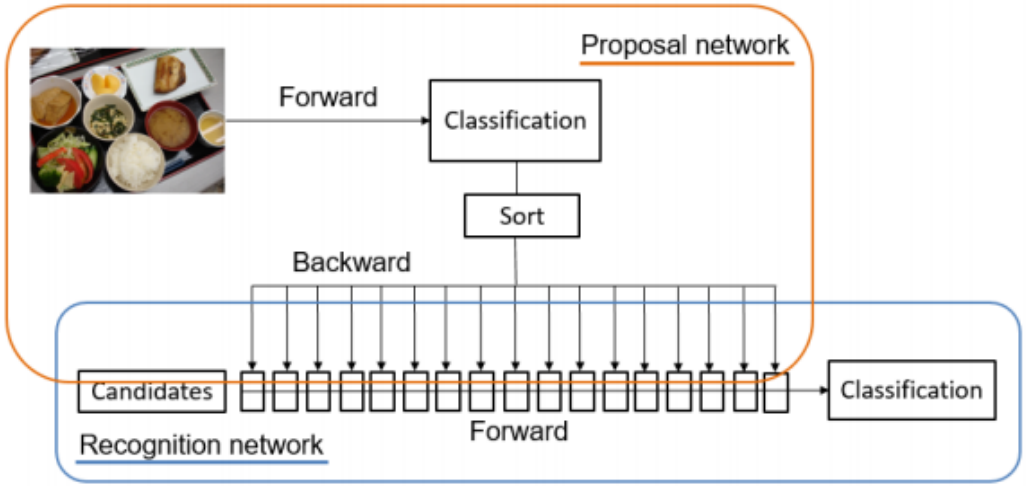
トースト　味噌汁　筑前煮　煮魚

領域候補の生成

から揚げ　生姜焼き　角煮　ビーフステーキ　野菜炒め　カツ丼　チキンライス

手法の概要

- 画像の認識結果をソート
- 上位Nクラスの食事の位置を推定
- 位置推定結果を領域候補に
- 領域候補をそれぞれ認識
- NMSにより統合



食事プロポーザルの認識

- 一般物体の認識と食事認識の違い
  - 小領域も高いスコアになりやすい
  - テクスチャの認識に近い

一般物体

背景

食事

ご飯

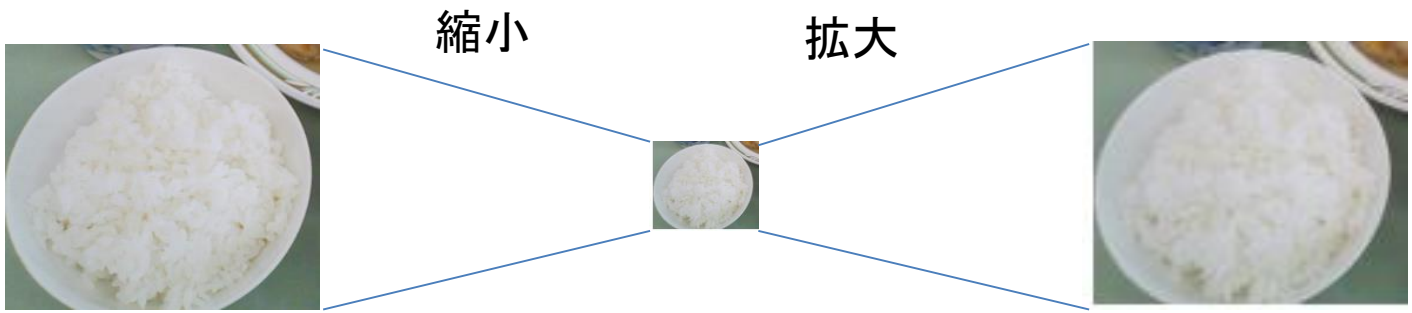
食事パッチ画像の学習

- 各食事について、中央付近からランダムに画像をクロップし、食事のパッチ画像を生成
  - パッチ画像が食事画像として認識されることを防ぐ
  - 各食品クラス＋各食品パッチクラスとして学習




低解像度画像の学習

- 低解像度の画像が食事のパッチとして分類されがちに
- 各画像について、低解像度画像を生成
  - 低解像度画像がパッチクラスとして認識されることを防ぐ



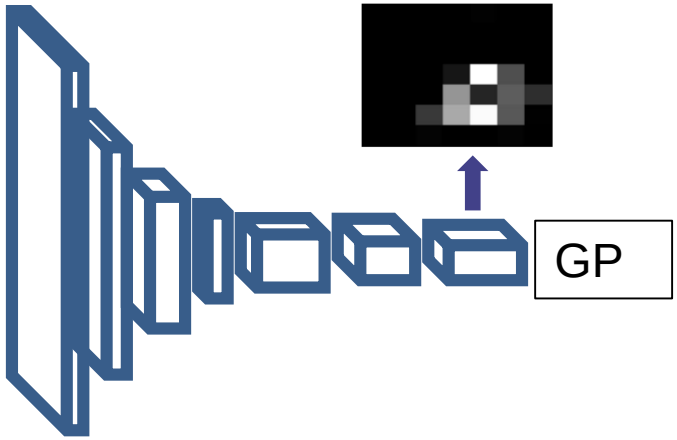
実験

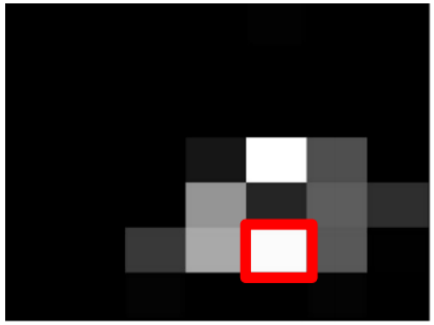
- 学習：
  - UECFOOD 100＋Webから収集した食事画像を用いて学習
    - 食事100カテゴリ：各1000枚　＋　非食事カテゴリ：10000枚
  - バウンディングボックスは用いずに学習
- テスト
  - UECFOOD 100 multiple food データセット
    - 100カテゴリの食事いずれかを含む複数品食事画像

領域候補の認識における検証

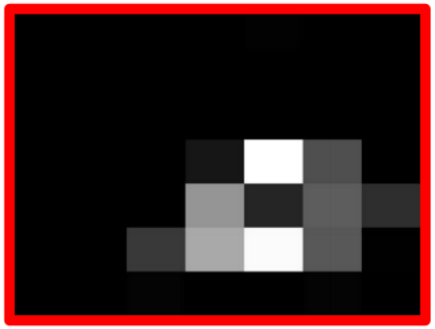
パッチ画像クラス	低解像度画像	100クラス（all）	53クラス（10枚以上）	11クラス（50枚以上）
―	―	33.5	35.1	33.3
○	―	32.2	34.8	31.8
○	○	**36.4**	**39.9**	**36.3**

Global Poolingの検証





Average pooling



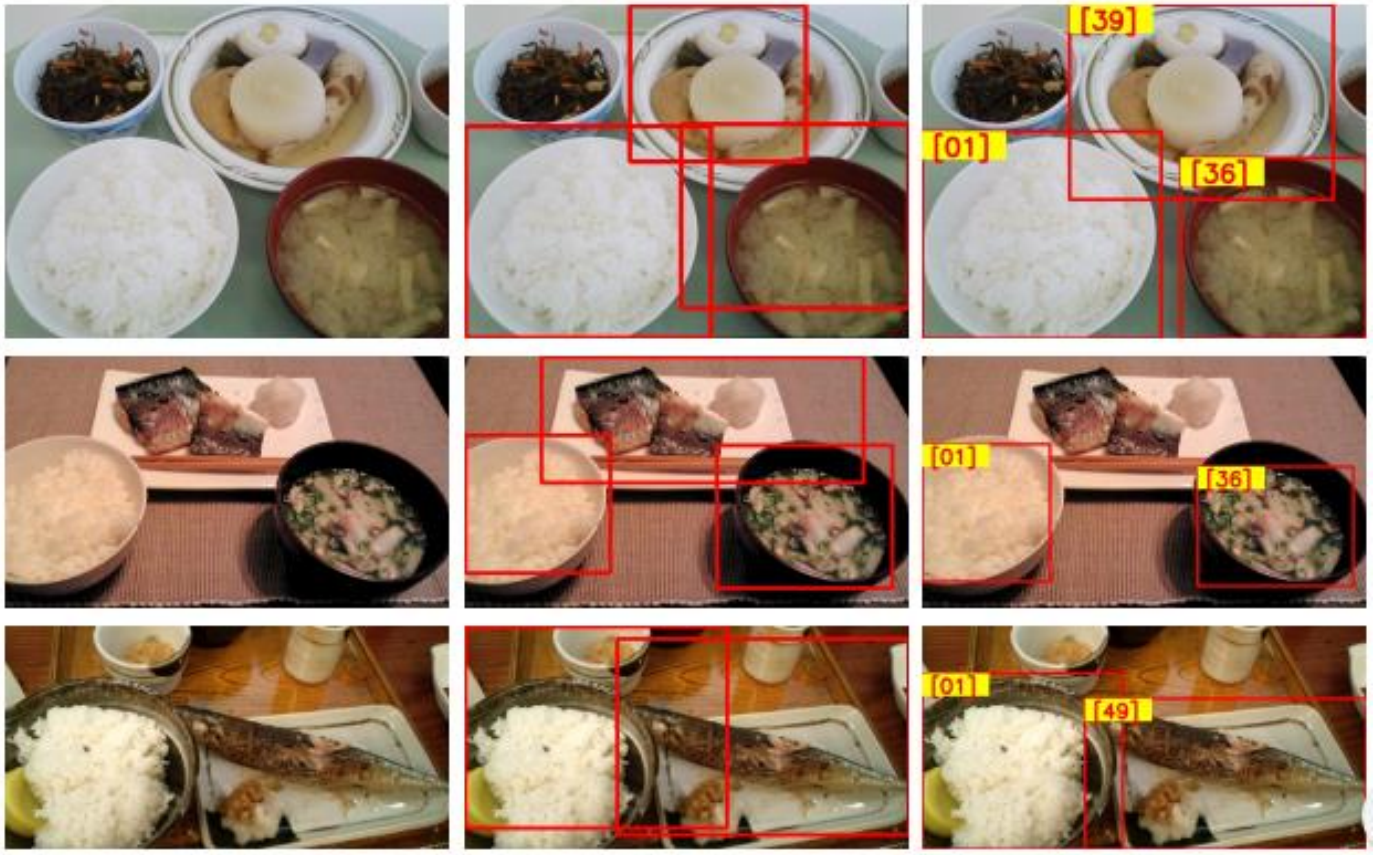
Max pooling

低解像度画像	100クラス（all）	53クラス（10枚以上）	11クラス（50枚以上）
Average pooling	36.4	39.9	36.3
Max pooling	**38.9**	**42.5**	**38.1**

他のプロポーザル手法との比較

手法	100クラス（all）	53クラス（10枚以上）	11クラス（50枚以上）	プロポーザル速度[s]	認識速度[s]
Selective Search	38.3	39.1	35.7	7.6	35.0
MCG	33.9	**43.7**	33.4	2.5	35.0
本手法　10 class	33.1	33.0	33.2	**0.5**	**1.1**
本手法　20 class	36.5	40.1	37.7	1.0	2.6
本手法　30 class	**38.9**	42.5	**38.1**	1.4	3.8

結果の例



結論

- シングルラベル画像のみを学習に用いて、弱教師あり物体検出を行った
- 既存のマルチラベルを想定した弱教師ありの手法を、プロポーザルによるアプローチへと応用した
- 既存のプロポーザル手法と比較して、高速かつ高精度な検出を行った